CSC2556

Lecture 11

Embedded EthiCS Module: Algorithmic Fairness

Agenda

- Introductions
- What is Embedded EthiCS? Why is it a good idea?
- Overview of algorithmic fairness
- Breakout activity 1
- Break
- Introduction to an application
- Breakout activity 2
- Hot topic discussion
- Conclusion

Introductions

Embedded EthiCS

- Ethical reasoning skill a must for computer scientists

- Embedded EthiCS™
  - Distributed pedagogy approach initiated at Harvard CS
  - Embedding ethical thinking and reasoning into CS courses

- Goals
  - To show CS students the extent to which ethical issues may arise when designing and deploying algorithms
  - To familiarize students with approaches to ethical design
  - To allow them to practice reasoning about ethics, articulating their positions, and incorporating their ideas into the systems they design

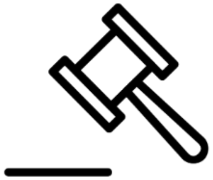
Today's Module: Algorithmic Fairness

Overview of Algorithmic Fairness

Algorithms Making Decisions

Loans

Bails



Self-Driving Cars

Ads

Hiring

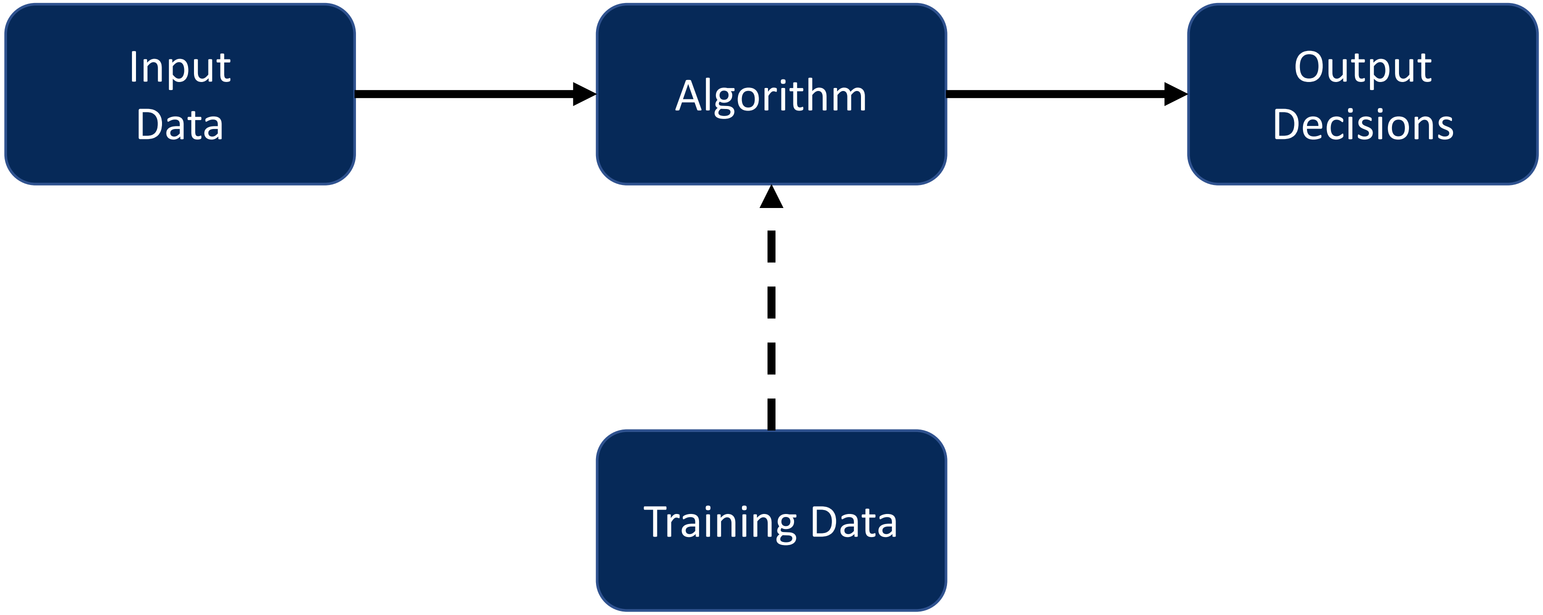
Organ Exchange

Examples of Unfairness

- COMPAS risk assessment for recidivism
- Resume screening tools which demonstrate bias against women and ethnic minorities
- Biased online ads
- Google translate following gender stereotypes
- Facial recognition technologies achieving dramatically different accuracy levels for different races

Turner Lee et al. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings, 2019.

Algorithms Making Decisions



Sources of Unfairness/Bias

- Bias in training/input data
  - Historical bias
  - Representation bias
  - Measurement bias
  - Simpson's paradox
  - …
- Bias in the algorithm
  - Direct discrimination
  - Indirect discrimination
  - Statistical discrimination
  - Justified vs unjustified discrimination
  - …

- We will mainly focus on bias in the algorithm
  - While social choice algorithms typically do not use training data, there can still be bias in input data

Mehrabi et al. A Survey on Bias and Fairness in Machine Learning. arXiv, 2019.

Important Terms

- Disparate Treatment
  - Individuals are treated differently because of animus against groups defined by race, gender, and other protected traits
  - [Equal Protection Clause of the 14th Amendment.]

- *Unjustified* Disparate Impact
  - A facially neutral policy produces disparate outcomes that are not justified by a legitimate, non-discriminatory interest.
  - [Civil Rights Act, Fair Housing Act, and various state statutes.]

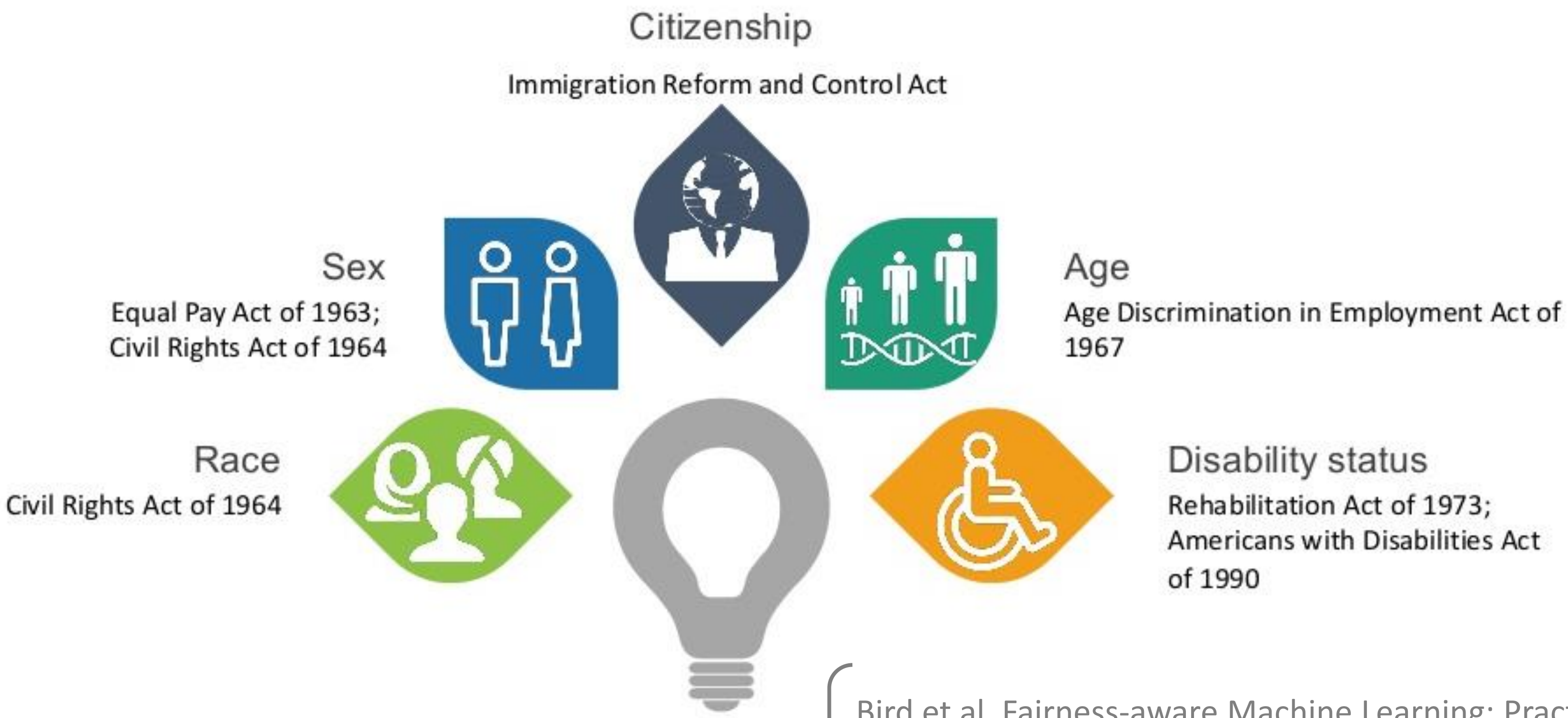
Corbett-Davies and Goel. Defining and Designing Fair Algorithms. EC Tutorial, 2018.

Types of Unfairness/Bias

- Outcome Fairness
  - Fairness in the outcomes produced by the algorithm
- Procedural fairness
  - Fairness of the algorithmic procedure
- In this course, we mainly focused on outcome fairness
  - We assumed that an agent's utility in a specific instance depends only on the outcome produced in *that* instance
  - But more generally, the utility may depend on the algorithm itself
  - Example: when I vote for candidate A and they lose, I may be unhappy, but may be more accepting of the outcome if I know that a fair rule like plurality was used to select the winner

Equal Entitlement

- Machine learning: protected attributes



Bird et al. Fairness-aware Machine Learning: Practical Challenges and Lessons Learned. KDD Tutorial, 2019.

Equal Entitlement

- Social choice
  - Agent neutrality
    - Permuting agent names permutes the outcome
  - Individual fairness notions with built-in equal entitlement
    - Proportionality
    - Envy-freeness
  - Sometimes agents may not be equally entitled
    - For example, a group of art collectors who wish to divide collectively bought artwork, but they contributed different amounts to the pool
    - Work on fairness notions with unequal entitlements

Definitions to Fairness

- Individual fairness
  - Individuals are treated fairly
- Group fairness (stronger than individual fairness)
  - Groups of individuals are treated fairly
- Group fairness (weaker than individual fairness)
  - *On average*, groups are treated fairly (but individuals members in those groups may be worse off)
- Extensions
  - Different entitlements, history, demographics, legal constraints, …

Economic Approaches

- Individual fairness
  - Proportionality: each individual gets their fair share
  - Envy-freeness: no individual envies another individual

- Group fairness
  - Core: each group of individuals gets their fair share
  - Group envy-freeness: no group envies another group
  - Stronger than individual fairness
  - There are also similar group fairness notions that are weaker than individual fairness

ML Approaches

- Popular fairness definitions
  - Demographic parity
  - Equal opportunity
  - Equalized odds
  - Calibration
  - Typically, pre-defined groups and binary outcomes
- Special cases of economic definitions
  - Restricted to the case of uniform preferences, e.g., everyone prefers the “+ve outcome” (e.g. receiving loan or bail) to the “-ve outcome”

Heidari et al. A Moral Framework for Understanding Fair ML through Economic Models of Equality of Opportunity. FAT*, 2019.

Hossain et al. Designing Fairly Fair Classifiers Via Economic Fairness Notions. TheWebConf, 2020.

Breakout Activity 1:
What does fairness entail?

Breakout Activity 1

- *What does fairness entail?*

  - You'll be divided into breakout groups

  - Each group will receive a hypothetical scenario, in which they will be tasked with making a decision that affects several entities

  - Each entity can be described with various features
    - E.g. a person can be described using their race, gender, education history, marital history, the number of attempts it took them to get their Ontario DL, whether they're afraid of heights, …

  - Most features would be *irrelevant* for the decision at hand

Breakout Activity 1

- Goals
  1. Identify the features which are relevant for the decision at hand
  2. Partition these features into two classes:
     - **Should Use:** For a good decision, one should take these features into account
     - **Must Avoid:** For fairness, the decision must not discriminate based on these features, as much as possible

- For example, a bank deciding whether to accept a loan application from an individual may consider "the number of previous loans defaulted" under *should use,* but race or gender under *must avoid*

Setup

1. You will be divided into 4 breakout groups
   - There are two scenarios, each will be assigned to two groups
2. An instructor/TA/ethics team member will join your breakout room and provide a Jamboard link
3. The first page of Jamboard will describe the scenario
   - Take a few minutes to read it carefully
4. Discuss with your group members
5. After the discussion, each group member separately adds stickies on 2nd page indicating features under “should use” and “must avoid”
   - Optionally include your initials
6. After the activity, we’ll compare the results

Synthesizing Thoughts from Activity 1: What does fairness entail?

Break!

Introduction to Participatory Budgeting

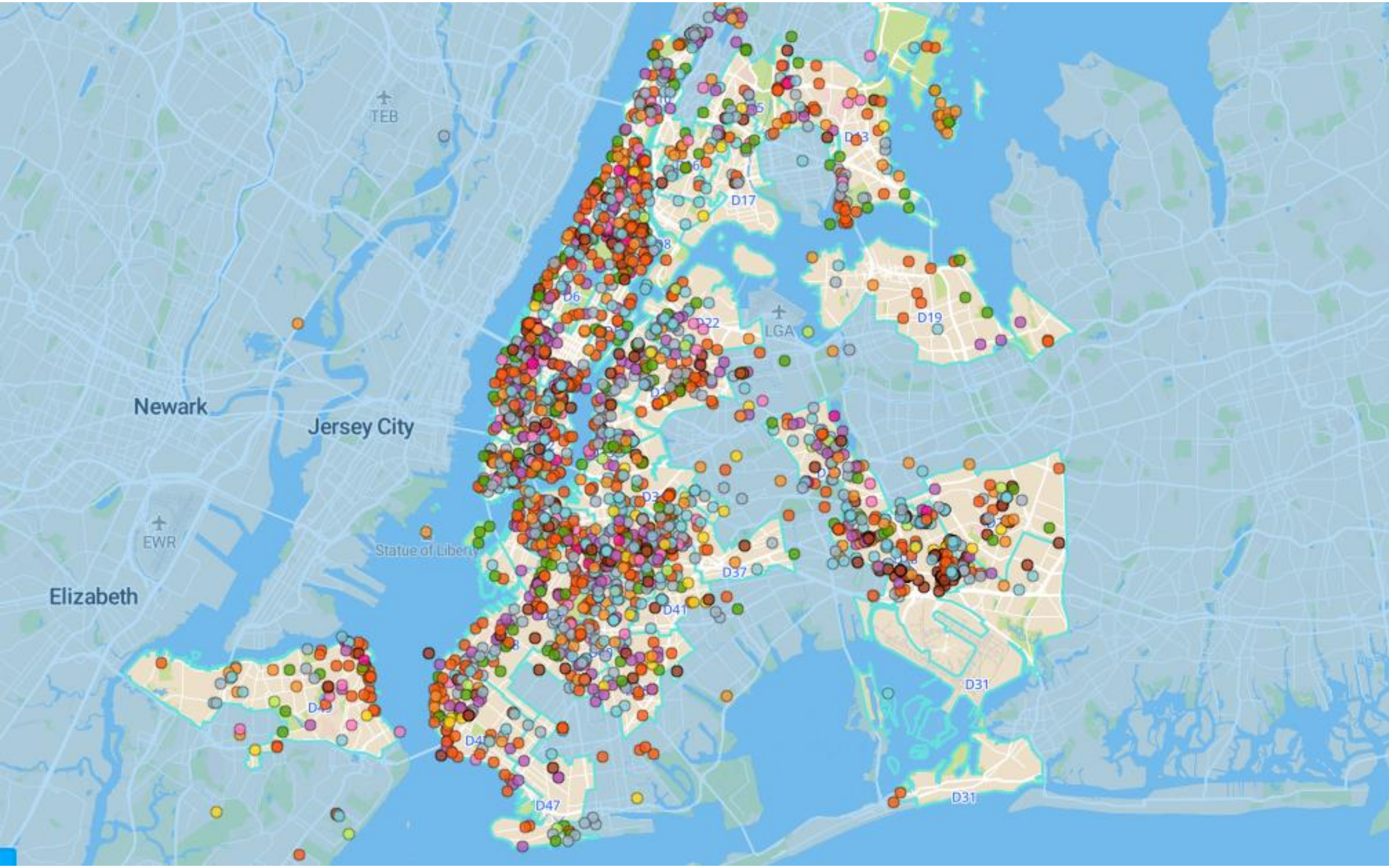
Participatory Budgeting

- Setting
  - Infrastructure projects proposed across a city
    - Each project p has a cost c_p
  - Budget B reserved by the city for funding these projects
    - A subset of projects S can be funded if $\sum_{p \in S} c_p \leq B$
  - Residents vote over the proposed projects
    - E.g. they could be asked to…
      - Select the top 3 projects they like (3-approval)
      - Rank the projects by how much they like them (ranking)
      - Rank the projects by “value-for-money” (VFM)
      - Select the best subset of projects according to them which fits the budget B (knapsack)

Participatory Budgeting

- Real-world application

- Hundreds of millions of dollars allocated each year worldwide
  - Paris ($100M/year), Boston, Cambridge, New York, San Francisco, …
  - Toronto (2015-2017), Toronto Community Housing (2001-present), Kitchener, …


TEB
Newark
Jersey City
EWR
Elizabeth
LGA
D6
D13
D17
D19
D22
D37
D41
D31
D31
D47

Project Examples

- Examples of real projects from Cambridge, PA

  - **Projects for healthy and safe recreation at our children's schools ($61,000)**
    - Field construction, synthetic turf, goal posts & installation for 25'x70' soccer field on east side of school.

  - **Remodel the Kitchen at the Youth Center ($200,000)**
    - The kitchen area in the Youth Center is in dire need of renovating. Replace the stove, dishwasher, cabinets, and countertops in the Frisoli Youth Center kitchen.

  - **Planting trees in the city ($119,400)**
    - Street trees cool the city, absorb pollution, & make our neighborhoods more livable! planting 100 new trees & building tree wells in the areas that need them most.

Goals

- Many goals not related to the final decision-making
  - Ensuring participation by diverse communities
  - Facilitating community discussion for filtering projects and to ensure an informed vote later on
  - …
- Final decision-making should balance the allocation of funds between…
  - Preferences of different sub-communities
  - Geographical regions
  - Category of projects (education, healthcare, parks, roads, …)
  - Low-cost versus high-cost projects
  - …

Approaches to PB

- Welfare maximization
  - Elicit or estimate the happiness of the community from each project
  - Select a feasible subset of projects maximizing the total happiness
  - For example, if each resident votes for their top 3 projects, select a feasible subset of projects to maximize the total number of votes

- Fairness: the core
  - Out of all residents N, there should be no $S \subseteq N$ such that by using their proportional share of the budget $B \cdot {}^{|S|}/_{|N|}$, they could fund a subset of projects which would make each of them happier than under the current decision

Breakout Activity 2:
How should the
public budget be allocated?

Breakout Activity 2

- Like before, you will be divided into breakout 4 groups
- An instructor/TA/ethics team member will join your breakout room and provide a Jamboard link
- The 1st page will describe a hypothetical PB scenario
  - Projects on an artificial map with their descriptions and costs
  - Total budget
  - Votes of the residents over the projects
- Read it carefully, discuss with your group members which subset of projects should be selected given the available information
- On the 2nd page, write down one or more proposed solutions

Synthesizing Thoughts from Activity 2: How should the public budget be allocated?

Algorithms vs Humans

- Arguments for algorithmic decision-making
  - *Potential to outperform humans in terms of accuracy and fairness*
    - They can leverage more data and potentially limitless computational power
  - *Potential to often be more transparent than humans*
    - Even if decisions are made using a black-box ML algorithm, being able to query the decisions in hypothetical scenarios makes it easier to assess fairness
  - *Potential to engage in deep mathematical reasoning about fairness*
    - Sometimes finding a fair outcome is an NP-hard problem
  - *Less bureaucracy, freeing up human time for other activities*
  - …

Algorithms vs Humans

- Arguments against algorithmic decision-making
  - *Algorithm may be designed to optimize the wrong objectives*
    - E.g. a social media platform designed to maximize the number of clicks rather than meaningful social connections, optimizing short-term objectives versus long-term goals
  - *Algorithms can often be less transparent than humans*
    - A black-box ML algorithm can be less transparent than a human following a well-documented and simple decision-making rule
  - *Being bound by a mathematical definition of fairness can be harmful*
    - No single definition may capture all facets of fairness in a context
  - *Potentially high energy consumption, impact on climate*
  - *…*

Algorithms vs Humans

- Poll 1
  - Suppose you are the mayor of Utopia City
  - Having heard of the amazing success of PB, you wish to conduct one
  - *If there are any complaints, you will be held accountable*
  - You have to choose between three systems for decision-making:
    1. A black-box machine learning algorithm, which can be trained to optimize any mathematically well-defined objectives
    2. A committee of city officials
    3. A committee of residents (citizen's assembly)
  - All three systems will try to optimize the same high-level goals and neither is fully transparent
  - Which system would you choose?

Algorithms vs Humans

- Poll 2
  - Consider the same problem, but now you're a resident of Utopia City
  - *You want to make sure that your voice is heard, the funds are allocated fairly and efficiently, and your neighborhood gets its deserved share of the funding*
  - You are given the option to provide your preference between the same three systems:
    1. A black-box machine learning algorithm, which can be trained to optimize any mathematically well-defined objectives
    2. A committee of city officials
    3. A committee of residents (citizen's assembly)
  - Again, all three systems will try to optimize the same high-level goals and neither is fully transparent
  - Which system would you prefer?

Concluding Remarks

- Improving algorithmic decision-making systems
  - Improving the quality and diversity of data sources
  - Causal inferences to determine which factors truly affect the decision at hand
  - Regulations and audits
  - Ensuring diverse ideas are represented within the designers of algorithmic decision-making systems

- Future challenges
  - Using algorithms to aid and improve human decision-making
    - E.g., matching reviewers to papers in conference reviewing
    - Also, other ways to mix human and algorithmic decision-making
  - Real-time ethical decision-making, e.g., in self-driving cars
  - …